

Maximum Likelihood Estimation for Finite Mixtures of Canonical Fundamental Skew t -Distributions: the Unification of the Unrestricted and Restricted Skew t -Mixture Models

Sharon X. Lee, Geoffrey J. McLachlan

Department of mathematics, the University of Queensland,
Brisbane, Australia

Abstract

In this paper, we present an algorithm for the fitting of a location-scale variant of the canonical fundamental skew t (CFUST) distribution, a superclass of the restricted and unrestricted skew t -distributions. In recent years, a few versions of the multivariate skew t (MST) model have been put forward, together with various EM-type algorithms for parameter estimation. These formulations adopted either a restricted or unrestricted characterization for their MST densities.

In this paper, we examine a natural generalization of these developments, employing the CFUST distribution as the parametric family for the component distributions, and point out that the restricted and unrestricted characterizations can be unified under this general formulation. We show that an exact implementation of the EM algorithm can be achieved for the CFUST distribution and mixtures of this distribution, and present some new analytical results for a conditional expectation involved in the E-step.

1 Introduction

To date, various parametric families of densities have been proposed in the literature to represent the multivariate skew t (MST) model, all having either a restricted or unrestricted characterization according to the classification scheme presented in Lee and McLachlan (2013b). In particular, the restricted MST (rMST) mixture model (or its equivalent variants by reparameterization) have been examined in Pyne et al. (2009), Cabral et al. (2012) and Vrbik and McNicholas (2012), while Lin (2010) and Lee and McLachlan (2011) focused on the unrestricted MST (uMST) mixture model adopting the characterization proposed by Sahu et al. (2003). In addition, Murray et al. (2013) recently considered an alternative construction that also adopted the term ‘skew t -distribution’, although referring to a quite different formulation more commonly known as the generalized-hyperbolic skew t (GHST) mixture model. Although this latter distribution is a restricted type of skew distribution (under the argument that the skewing variable is a scalar), it differs from the rMST distribution in a number ways, such as having different behaviour in its tails and not becoming a skew normal distribution as a limiting and/or special case.

In this paper, we consider a natural and straightforward extension of the uMST distribution, where the skewing function of its density is given by a q -dimensional t -distribution function,

and the skewness parameter is a general p by q matrix. This formulation coincides with a special case of the unified skew t (SUT) distribution (Arellano-Valle and Azzalini, 2006) and the fundamental skew t (FUST) distribution (Arellano-Valle and Genton, 2005), and is known as a location-scale variant of the canonical fundamental skew t (CFUST) distribution. This characterization includes the restricted form of the MST distribution.

As discussed in Lee and McLachlan (2013a), while exact implementation of the Expectation-Maximization (EM) algorithm have been achievable for restricted characterizations of the component skew t -distributions, parameter estimation for the unrestricted and more general characterizations is substantially more intricate. In previous work on mixtures of unrestricted skew t -distributions using Sahu et al. (2003)'s characterization (a submodel of the CFUST distribution considered in this paper), Lin (2010) resorted to Monte Carlo methods for the computation of intractable conditional expectations involved in the E-step, while Lee and McLachlan (2013a) adopted the OSL (one-step late) approach to evaluating one of these difficult conditional expectation. In contrast for the rMST distribution, a number of closed-form implementations of the EM algorithm have been put forward (see Lee and McLachlan (2013a) for further discussion).

This paper presents an EM fitting algorithm for the FM-CFUST model. Under this general formulation, the aforementioned restricted and unrestricted skew t -mixture models can be obtained as special cases and, as such, the expressions on the E- and M-steps of the proposed algorithm would apply directly in these situations. In contrast to the approach of Lin (2010) which relies on computationally intensive numerical approximations, our development was motivated by the desire for an exact implementation of the EM algorithm using the truncated moments approach (Lee and McLachlan, 2013a, Ho et al., 2012), leading to much faster and more accurate results. Also, we derive a new exact expression for the remaining E-step conditional expectation that had previously relied on approximation or numerical methods.

2 The CFUST distribution

The multivariate canonical fundamental skew t -distribution (CFUST) was introduced by Arellano-Valle and Genton (2005). A location-scale variant of the CFUST distribution can be characterized as follows. Let \mathbf{U}_1 be p -dimensional random vector and \mathbf{U}_0 a q -dimensional random vector. Suppose that, conditional on a scalar gamma variable $w \sim \text{gamma}(\frac{\nu}{2}, \frac{\nu}{2})$, the joint distribution of \mathbf{U}_0 and \mathbf{U}_1 is given by

$$\begin{bmatrix} \mathbf{U}_0 \\ \mathbf{U}_1 \end{bmatrix} \sim N_{q+p} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \frac{1}{w} \begin{bmatrix} \mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} \end{bmatrix} \right), \quad (1)$$

where ν is a scalar, \mathbf{I}_q denotes the $q \times q$ identity matrix, $\mathbf{\Sigma}$ is a positive definite scale matrix, and $\mathbf{0}$ is a vector/matrix of zeros with appropriate dimensions. Then

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Delta} |\mathbf{U}_0| + \mathbf{U}_1 \quad (2)$$

follows the CFUST distribution, whose density is defined by

$$f(\mathbf{y}; \boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\Delta}, \nu) = 2^q t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Omega}, \nu) T_q \left(\mathbf{q}(\mathbf{y}) \sqrt{\frac{\nu + p}{\nu + d(\mathbf{y})}}; \mathbf{0}, \boldsymbol{\Lambda}, \nu + p \right), \quad (3)$$

where

$$\begin{aligned} \boldsymbol{\Omega} &= \mathbf{\Sigma} + \boldsymbol{\Delta} \boldsymbol{\Delta}^T, \\ \mathbf{q} &= \boldsymbol{\Delta}^T \boldsymbol{\Omega}^{-1} (\mathbf{y} - \boldsymbol{\mu}), \\ \boldsymbol{\Lambda} &= \mathbf{I}_p - \boldsymbol{\Delta}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\Delta}, \\ d(\mathbf{y}) &= (\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Omega}^{-1} (\mathbf{y} - \boldsymbol{\mu}). \end{aligned}$$

Here, we let $t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Omega}, \nu)$ denotes the p -dimensional t -distribution with location parameter $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Omega}$, and degree of freedom $\nu + p$, and $T_q(\cdot)$ is the q -dimensional cumulative t -distribution function. In the above, $|\mathbf{U}_0|$ denotes the vector whose i th element is the magnitude of the i th element of the vector \mathbf{U}_0 .

2.1 The unrestricted characterization as a special case

As mentioned previously, the characterization of Sahu et al. (2003) corresponds to the CFUST distribution with $q = p$ and $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\delta})$. Hence, we can write $\boldsymbol{\Lambda} = \mathbf{I}_p - \boldsymbol{\Delta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}$, $\mathbf{q} = \boldsymbol{\Delta}\boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ and $\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{\Delta}^2$, and the uMST density is given by

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\delta}, \nu) = 2^p t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Omega}, \nu) T_p\left(\mathbf{q}(\mathbf{y}) \sqrt{\frac{\nu + p}{\nu + d(\mathbf{y})}}; \mathbf{0}, \boldsymbol{\Lambda}, \nu + p\right). \quad (4)$$

The uMST distribution has the same stochastic representation as given by (2) and (1) with q replaced by p .

2.2 The restricted characterization as a special case

In the restricted characterization of the MST distribution, the convolution-type stochastic representation is given by

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\delta} |U_0| + \mathbf{U}_1 \quad (5)$$

where

$$\begin{bmatrix} U_0 \\ \mathbf{U}_1 \end{bmatrix} \sim N_{1+p}\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \frac{1}{w} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \end{bmatrix}\right). \quad (6)$$

It follows that the density of the rMST distribution is given by

$$f(\mathbf{y}) = 2t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Omega}, \nu) T_1\left(\boldsymbol{\delta}^T \boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu}), \left(\frac{\nu + d(\mathbf{y})}{\nu + p}\right) (1 - \boldsymbol{\delta}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\delta}), \nu + p\right). \quad (7)$$

On directly comparing (5) and (2), it can be observed that the skewing effect of $\boldsymbol{\delta}|U_0|$ in the rMST distribution can be achieved by the CFUST formulation in a number of ways, including

- taking $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\delta})$ and in the unrestricted case, and setting $\mathbf{U}_0 = U_0 \mathbf{1}_p$ in (2), where $\mathbf{1}_p$ denotes a p -dimensional vector with elements one;
- constraining the skewness matrix $\boldsymbol{\Delta}$ to be a $p \times p$ zero matrix except for one column which is given by $\boldsymbol{\delta}$, and setting the corresponding element in \mathbf{U}_0 to be U_0 ; or
- setting $q = 1$.

As a corollary, it can be seen that the special case of $q = p$ of the CFUST distribution encompass both the rMST and uMST distribution through structural constraints on the skewness parameter. This is given by taking

$$\boldsymbol{\Delta} = \begin{bmatrix} \delta_1 & 0 & \dots & 0 \\ \delta_2 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \delta_p & 0 & \dots & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Delta} = \begin{bmatrix} \delta_1 & 0 & \dots & 0 \\ 0 & \delta_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \delta_p \end{bmatrix} \quad (8)$$

for the rMST and uMST distributions respectively, where $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_p)^T$. Note that for the rMST case, $\boldsymbol{\delta}$ can be placed in any one of the columns of $\boldsymbol{\Delta}$, not necessarily the first column as given in (8).

3 Parameter estimation via the EM algorithm

We are interested in the finite mixtures with component densities given by (3). A g -component mixture of CFUST distributions has density

$$f(\mathbf{y}; \Psi) = \sum_{h=1}^g \pi_h f(\mathbf{y}; \boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h, \boldsymbol{\Delta}_h, \nu_h), \quad (9)$$

where π_h ($h = 1, \dots, g$) are the mixing proportions and $f(\cdot)$ denotes a CFUST density given by (3). We shall adopt the acronym FM-CFUST for (9). The vector $\Psi = (\pi_1, \dots, \pi_{g-1}, \boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_g^T)$ contains all the unknown parameters of the model, with $\boldsymbol{\theta}_h$ containing the elements of $\boldsymbol{\mu}_h$ and $\boldsymbol{\delta}_h$, the distinct elements of $\boldsymbol{\Sigma}_h$, and ν_h .

Similar to the rMST and uMST distributions, the CFUST admits a convenient hierarchical representation that greatly facilitates parameter estimation via the EM algorithm:

$$\begin{aligned} \mathbf{Y} \mid \mathbf{U}, w &\sim N_p\left(\boldsymbol{\mu} + \boldsymbol{\Delta}\mathbf{u}, \frac{1}{w}\boldsymbol{\Sigma}\right) \\ \mathbf{U} \mid w &\sim HN_q\left(\mathbf{0}, \frac{1}{w}\mathbf{I}_m\right) \\ w &\sim \text{gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \end{aligned} \quad (10)$$

where $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate normal distribution and $HN_q(0, \boldsymbol{\Sigma})$ represents the q -dimensional half-normal distribution with mean $\mathbf{0}$ and scale matrix $\boldsymbol{\Sigma}$. The EM algorithm for FM-CFUST then proceeds in a similar way as for the FM-uMST model as described in Lee and McLachlan (2013a).

3.1 E-Step

In the E-step, the following conditional expectations are computed:

$$\begin{aligned} z_{hj}^{(k)} &= E_{\Psi^{(k)}} [z_{hj} = 1 \mid \mathbf{y}_j] \\ w_{hj}^{(k)} &= E_{\Psi^{(k)}} [w_{hj} \mid \mathbf{y}_j, z_{hj} = 1], \\ e_{1hj}^{(k)} &= E_{\Psi^{(k)}} [\log(w_{hj}) \mid \mathbf{y}_j, z_{hj} = 1], \\ \mathbf{e}_{2hj}^{(k)} &= E_{\Psi^{(k)}} [w_{hj}\mathbf{u}_{hj} \mid \mathbf{y}_j, z_{hj} = 1], \\ \mathbf{e}_{3hj}^{(k)} &= E_{\Psi^{(k)}} [w_{hj}\mathbf{u}_{hj}\mathbf{u}_{hj}^T \mid \mathbf{y}_j, z_{hj} = 1]. \end{aligned}$$

They are given by

$$\begin{aligned}
z_{hj}^{(k)} &= \frac{\pi_h f(\mathbf{y}_j; \boldsymbol{\mu}_h^{(k)}, \boldsymbol{\Sigma}_h^{(k)}, \boldsymbol{\delta}_h^{(k)}, \nu_h^{(k)})}{f(\mathbf{y}_j; \boldsymbol{\Psi}^{(k)})}, \\
w_{hj}^{(k)} &= \left(\frac{\nu_h^{(k)} + p}{\binom{(k)}{h} + d_h^{(k)}(\mathbf{y}_j)} \right) \frac{T_q \left(\mathbf{q}_h^{(k)} \sqrt{\frac{\nu_h^{(k)} + p + 2}{\nu_h^{(k)} d_h^{(k)}(\mathbf{y}_j)}}; \mathbf{0}, \boldsymbol{\Lambda}_h^{(k)}, \nu_h^{(k)} + p + 2 \right)}{T_q \left(\mathbf{q}_h^{(k)} \sqrt{\frac{\nu_h^{(k)} + p}{\nu_h^{(k)} d_h^{(k)}(\mathbf{y}_j)}}; \mathbf{0}, \boldsymbol{\Lambda}_h^{(k)}, \nu_h^{(k)} + p \right)}, \\
e_{1hj}^{(k)} &= w_{hj}^{(k)} - \log \left(\frac{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)}{2} \right) + \left(\frac{\nu_h^{(k)} + p}{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)} \right) - \psi \left(\frac{\nu_h^{(k)} + p}{2} \right) \quad (11)
\end{aligned}$$

$$\begin{aligned}
\mathbf{e}_{2,hj}^{(k)} &= w_{hj}^{(k)} E_{\boldsymbol{\Psi}^{(k)}}[\mathbf{u}_{hj} \mid \mathbf{y}_j], \\
\mathbf{e}_{3hj}^{(k)} &= w_{hj}^{(k)} E_{\boldsymbol{\Psi}^{(k)}}[\mathbf{u}_{hj} \mathbf{u}_{hj}^T \mid \mathbf{y}_j], \quad (12)
\end{aligned}$$

where $\mathbf{U}_{hj} \mid \mathbf{y}_j$ has a q -dimensional truncated t -distribution given by

$$\mathbf{U}_{hj} \mid \mathbf{y}_j \sim tt_q \left(\mathbf{q}_h^{(k)}, \left(\frac{\nu_h^{(k)} + d_h(\mathbf{y}_j)}{\nu_h^{(k)} + p + 2} \right) \boldsymbol{\Lambda}_h^{(k)}, \nu_h^{(k)} + p + 2; \mathbb{R}^+ \right). \quad (13)$$

Note that (11) is obtained using a OSL EM algorithm. The derivations for these results are analogous to Lee and McLachlan (2013a). Alternatively, a series truncation approach can be used in place of (11), which exploits an exact representation of this conditional expectation, will be presented later.

3.2 M-Step

On the $(k+1)$ th M-step, the model parameters are updated with

$$\begin{aligned}
\pi_h^{(k+1)} &= \frac{1}{n} \sum_{j=1}^n z_{hj}^{(k)}, \\
\boldsymbol{\mu}_h^{(k+1)} &= \frac{\sum_{j=1}^n z_{hj} w_{hj}^{(k)} \mathbf{y}_j - \boldsymbol{\Delta}_h^{(k)} \sum_{j=1}^n z_{hj}^{(k)} \mathbf{e}_{2hj}^{(k)}}{\sum_{j=1}^n z_{hj}^{(k)} w_{hj}^{(k)}}, \\
\boldsymbol{\Delta}_h^{(k+1)} &= \left[\sum_{j=1}^n z_{hj}^{(k)} \left(\mathbf{y}_j - \boldsymbol{\mu}_h^{(k+1)} \right) \mathbf{e}_{2hj}^{(k)T} \right] \left[\sum_{j=1}^n z_{hj}^{(k)} \mathbf{e}_{3hj}^{(k)} \right]^{-1}, \quad (14)
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\Sigma}_h^{(k+1)} &= \left[\sum_{j=1}^n z_{hj}^{(k)} \right]^{-1} \left\{ \sum_{j=1}^n z_{hj}^{(k)} \left[w_{hj}^{(k)} \left(\mathbf{y}_j - \boldsymbol{\mu}_h^{(k+1)} \right) \left(\mathbf{y}_j - \boldsymbol{\mu}_h^{(k+1)} \right)^T - \boldsymbol{\Delta}_h^{(k+1)} \mathbf{e}_{2hj}^{(k)} \left(\mathbf{y}_j - \boldsymbol{\mu}_h^{(k+1)} \right)^T \right. \right. \\
&\quad \left. \left. - \left(\mathbf{y}_j - \boldsymbol{\mu}_h^{(k+1)} \right) \mathbf{e}_{2hj}^{(k)T} \boldsymbol{\Delta}_h^{(k+1)T} + \boldsymbol{\Delta}_h^{(k+1)} \mathbf{e}_{3hj}^{(k)T} \boldsymbol{\Delta}_h^{(k+1)T} \right] \right\}. \quad (15)
\end{aligned}$$

An update of the degrees of freedom is obtained by solving for ν_h :

$$0 = \left(\sum_{h=1}^n z_{hj}^{(k)} \right) \left[\log \left(\frac{\nu_h}{2} \right) - \psi \left(\frac{\nu_h}{2} \right) + 1 \right] - \left(\sum_{j=1}^n e_{1hj}^{(k)} - w_{hj}^{(k)} \right).$$

3.3 The rMST and uMST as special cases

As pointed out previously, the rMST and uMST distributions can be obtained as special cases of the CFUST distribution by imposing structural constraints on the skewness parameter Δ . In the case of the uMST distribution, the diagonal elements of the component skewness matrices are estimated as

$$\delta_h^{(k+1)} = \left(\Sigma_h^{(k)-1} \circ \sum_{j=1}^n \tau_{hj}^{(k)} e_{4,hj}^{(k)} \right)^{-1} \text{diag} \left[\Sigma_h^{(k)-1} \sum_{j=1}^n \tau_{hj}^{(k)} (\mathbf{y}_j - \boldsymbol{\mu}_h^{(k+1)}) e_{3,hj}^{(k)T} \right], \quad (16)$$

where \circ denotes the elementwise product.

Under the constraint that Δ_h is zero except for the first column δ_h , that is, $\Delta = [\delta \mathbf{0} \dots \mathbf{0}]$, (14) reduces to estimating

$$\delta_h^{(k+1)} = \frac{\sum_{j=1}^n \tau_{hj}^{(k)} [e_{3,hj}^{(k)}]_1 (\mathbf{y}_j - \boldsymbol{\mu}_h^{(k+1)})}{\sum_{j=1}^n \tau_{hj}^{(k)} [e_{4,hj}^{(k)}]_{11}}, \quad (17)$$

where $[e_{3,hj}^{(k)}]_1$ and $[e_{4,hj}^{(k)}]_{11}$ denote the first element of $e_{3,hj}^{(k)}$ and $e_{4,hj}^{(k)}$, respectively. On comparing the above with the corresponding expression for the rMST mixture model (see, for example, equation (9) from Wang et al. (2009)), it can be observed that the two expressions are indeed equivalent.

3.4 New results on $e_{1,hj}^{(k)}$: the truncated series approach

We now present the derivation of a new exact expression for $e_{1,hj}^{(k)}$, expressed in terms of an infinite series involving multivariate t -distribution functions. By the law of iterated expectations,

$$\begin{aligned} e_{1,hj}^{(k)} &= E_{\Psi^{(k)}} [\log(W_j) \mid \mathbf{y}_j] \\ &= E_{\Psi^{(k)}} [E_{\Psi^{(k)}} [\log(W_j) \mid \mathbf{u}_j, \mathbf{y}_j] \mid \mathbf{y}_j] \\ &= E_{\Psi^{(k)}} \left[\psi \left(\frac{\nu_h^{(k)} + p}{2} \right) - \log \left(\frac{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j) + (\mathbf{u}_j - \mathbf{q}_{hj}^{(k)})^T \Lambda_h^{(k)-1} (\mathbf{u}_j - \mathbf{q}_{hj}^{(k)})}{2} \right) \mid \mathbf{y}_j \right] \\ &= \psi \left(\frac{\nu_h^{(k)} + p}{2} \right) - \log \left(\frac{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)}{2} \right) - E_{\Psi^{(k)}} \left[\log \left(1 + \frac{(\mathbf{u}_j - \mathbf{q}_{hj}^{(k)})^T \Lambda_h^{(k)-1} (\mathbf{u}_j - \mathbf{q}_{hj}^{(k)})}{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)} \right) \right]. \end{aligned} \quad (18)$$

To calculate the last term in (18), note that the conditional density of \mathbf{u}_j given \mathbf{y}_j can be written as

$$f(\mathbf{u}_j \mid \mathbf{y}_j) = \frac{t_q \left(\mathbf{u}_j; \mathbf{q}_{hj}^{(k)}, \left(\frac{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)}{\nu_h^{(k)} + p} \right) \Lambda_h^{(k)}, \nu_h^{(k)} + p \right)}{T_q \left(\mathbf{q}_{hj}^{(k)}; \mathbf{0}, \left(\frac{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)}{\nu_h^{(k)} + p} \right) \Lambda_h^{(k)}, \nu_h^{(k)} + p \right)}. \quad (19)$$

After some algebraic manipulations, it follows that the expectation in (18) can be reduced to a simple closed-form expression (see equation (76) from Lee and McLachlan (2013a)), except

for a term involving an integral given by

$$\begin{aligned}
S_{1,hj}^{(k)} &= \int_{\mathbf{0}}^{\infty} \log \left(1 + \frac{(\mathbf{u}_j - \mathbf{q}_{hj}^{(k)})^T \mathbf{\Lambda}_h^{(k)-1} (\mathbf{u}_j - \mathbf{q}_{hj}^{(k)})}{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)} \right) \\
&\quad t_q \left(\mathbf{u}_j; \mathbf{q}_{hj}^{(k)}, \left(\frac{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)}{\nu_h^{(k)} + p} \right) \mathbf{\Lambda}_h^{(k)}, \nu_h^{(k)} + p \right) d\mathbf{u}_j \\
&= \sum_{r=1}^{\infty} \sum_{s=0}^r \frac{(-1)^{2r-s-1}}{r} \binom{r}{s} \int_{\mathbf{0}}^{\infty} \left(1 + \frac{(\mathbf{u}_j - \mathbf{q}_{hj}^{(k)})^T \mathbf{\Lambda}_h^{(k)-1} (\mathbf{u}_j - \mathbf{q}_{hj}^{(k)})}{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)} \right)^{-s} \\
&\quad t_q \left(\mathbf{u}_j; \mathbf{q}_{hj}^{(k)}, \left(\frac{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)}{\nu_h^{(k)} + p} \right) \mathbf{\Lambda}_h^{(k)}, \nu_h^{(k)} + p \right) d\mathbf{u}_j \tag{20}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{r=1}^{\infty} \sum_{s=0}^r \frac{(-1)^{2r-s-1}}{r} \binom{r}{s} \frac{\Gamma \left(\frac{\nu_h^{(s)} + p}{2} + s \right)}{\Gamma \left(\frac{\nu_h^{(k)} + p}{2} \right)} \frac{\Gamma \left(\frac{\nu_h^{(k)} + p + q}{2} \right)}{\Gamma \left(\frac{\nu_h^{(k)} + p + q}{2} + s \right)} \\
&\quad \int_{\mathbf{0}}^{\infty} t_q \left(\mathbf{u}_j; \mathbf{q}_{hj}^{(k)}, \left(\frac{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)}{\nu_h^{(k)} + p + 2s} \right) \mathbf{\Lambda}_h^{(k)}, \nu_h^{(k)} + p + 2s \right) d\mathbf{u}_j \tag{21}
\end{aligned}$$

$$\begin{aligned}
&= \frac{\Gamma \left(\frac{\nu_h^{(k)} + p + q}{2} \right)}{\Gamma \left(\frac{\nu_h^{(k)} + p}{2} \right)} \sum_{r=1}^{\infty} \sum_{s=0}^r \frac{(-1)^{2r-s-1}}{r} \binom{r}{s} \frac{\Gamma \left(\frac{\nu_h^{(s)} + p}{2} + s \right)}{\Gamma \left(\frac{\nu_h^{(k)} + p + q}{2} + s \right)} \\
&\quad T_q \left(\mathbf{q}_{hj}^{(k)}; \mathbf{0}, \left(\frac{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)}{\nu_h^{(k)} + p + 2s} \right) \mathbf{\Lambda}_h^{(k)}, \nu_h^{(k)} + p + 2s \right) \tag{22}
\end{aligned}$$

In (20) above, we have applied a Taylor expansion of $\log(x)$ for $0 < x < 2$ and the binomial expansion formula to obtain a series expression for $\log(x)$, given by

$$\log(x) = \sum_{r=1}^{\infty} \sum_{s=0}^r \frac{(-1)^{2r-s-1}}{r} \binom{r}{s} x^s, \quad \text{for } 0 < x < 2.$$

Note that the term inside the logarithm in (21) is always greater than one. Hence, its inverse lies between 0 and 1, and we can apply the above series representation. In practice, $S_{1,hj}^{(k)}$ can be approximated by a (small) finite number of terms.

It follows that $e_{1,hj}^{(k)}$ is given by

$$\begin{aligned}
e_{1,hj}^{(k)} &= \psi\left(\frac{\nu_h^{(k)} + p}{2}\right) - \log\left(\frac{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)}{2}\right) \\
&\quad + \frac{\Gamma\left(\frac{\nu_h^{(k)} + p + q}{2}\right)}{\Gamma\left(\frac{\nu_h^{(k)} + p}{2}\right)} T_q^{-1}\left(\mathbf{q}_{hj}^{(k)}; \mathbf{0}, \left(\frac{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)}{\nu_h^{(k)} + p}\right) \mathbf{\Lambda}_h^{(k)}, \nu_h^{(k)} + p\right) \\
&\quad \sum_{r=1}^{\infty} \sum_{s=0}^r \frac{(-1)^{2r-s-1}}{r} \binom{r}{s} \frac{\Gamma\left(\frac{\nu_h^{(s)} + p}{2} + s\right)}{\Gamma\left(\frac{\nu_h^{(k)} + p + q}{2} + s\right)} \\
&\quad T_q\left(\mathbf{q}_{hj}^{(k)}; \mathbf{0}, \left(\frac{\nu_h^{(k)} + d_h^{(k)}(\mathbf{y}_j)}{\nu_h^{(k)} + p + 2s}\right) \mathbf{\Lambda}_h^{(k)}, \nu_h^{(k)} + p + 2s\right). \tag{23}
\end{aligned}$$

4 Conclusions

We have introduced a finite mixture of CFUST distributions, where the component densities are the generalization of the restricted and unrestricted skew t -distributions. Parameter estimation via the EM algorithm has been outlined, following the truncated moments approach in Ho et al. (2012) and Lee and McLachlan (2013a). Furthermore, new results have been derived for conditional expectation $w_{hj}^{(k)}$, exploiting an infinite series representation of the logarithm function. This leads to a full implementation of the EM algorithm for the FM-CFUST model with exact and closed-form expressions for all E-step conditional expectations. Simulations and applications of the FM-CFUST model, as well as extension to factor analyzers, will be treated in a forthcoming manuscript.

References

- Arellano-Valle RB, Azzalini A (2006) On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics* 33:561–574
- Arellano-Valle RB, Genton MG (2005) On fundamental skew distributions. *Journal of Multivariate Analysis* 96:93–116
- Cabral CS, Lachos VH, Prates MO (2012) Multivariate mixture modeling using skew-normal independent distributions. *Computational Statistics and Data Analysis* 56:126–142
- Ho HJ, Lin TI, Chen HY, Wang WL (2012) Some results on the truncated multivariate t distribution. *Journal of Statistical Planning and Inference* 142:25–40
- Lee S, McLachlan GJ (2011) On the fitting of mixtures of multivariate skew t -distributions via the EM algorithm. arXiv:11094706 [statME]
- Lee S, McLachlan GJ (2013a) Finite mixtures of multivariate skew t -distributions: some recent and new results. *Statistics and Computing* DOI 10.1007/s11222-012-9362-4

- Lee SX, McLachlan GJ (2013b) On mixtures of skew-normal and skew t -distributions. *Advances in Data Analysis and Classification* 7:241–266
- Lin TI (2010) Robust mixture modeling using multivariate skew t distribution. *Statistics and Computing* 20:343–356
- Murray PM, Browne BP, McNicholas PD (2013) Mixtures of skew- t factor analyzers. Preprint arXiv: 13054301 [statME]
- Pyne S, Hu X, Wang K, Rossin E, Lin TI, Maier LM, Baecher-Allan C, McLachlan GJ, Tamayo P, Hafler DA, De Jager PL, Mesirov JP (2009) Automated high-dimensional flow cytometric data analysis. *Proceedings of the National Academy of Sciences USA* 106:8519–8524
- Sahu SK, Dey DK, Branco MD (2003) A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics* 31:129–150
- Vrbik I, McNicholas PD (2012) Analytic calculations for the EM algorithm for multivariate skew t -mixture models. *Statistics and Probability Letters* 82:1169–1174
- Wang K, Ng SK, McLachlan GJ (2009) Multivariate skew t mixture models: applications to fluorescence-activated cell sorting data. In: Shi H, Zhang Y, Bottema MJ, Lovell BC, Maeder AJ (eds) *DICTA 2009 (Conference of Digital Image Computing: Techniques and Applications, Melbourne)*, IEEE Computer Society, Los Alamitos, California, pp 526–531